Weekly Report

11/30/2015-12/06/2015

Research

In order to complete the final project(face recognition) of computer vision, I read some papers and articles about two methods: Eigenfaces and Fisherfaces.

Eigenfaces based on PCA

The Eigenfaces method use PCA to project the image space to a low dimensional feature space. Given a set of N sample images $\{x_1, x_2, ..., x_N\}$,$x_i \in \mathbb{R}^n$,$n = height * width$. If original n-dimensional images are projected into an m-dimensional space, the scatter across all images are supposed to be max. The feature vector $y_k \in \mathbb{R}^m$ of image x_k is defined as

$$y_k = W^T x_k$$

where $W^T \in \mathbb{R}^{n \times m} isaprojectionmatrix.$ And the total scatter matrix S_T is defined as

$$S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T$$

Then, the scatter matrix of feature vectors $\{y_1, y_2, ..., y_N\}$ is $W^T S_T W$. PCA always chooses a projection matrix W_{opt} to maximize the scatter matrix of projected images.

$$\begin{aligned} W_{opt} &= \arg\max |W^T S_T W| \\ &= [w_1 w_2 ... w_m] \end{aligned}$$

where $\{w_i | i = 1, 2, ..., m\}$is the set of n-dimensional eigenvectors of S_T corresponding to the m largest eigenvalues $\{\lambda_i | i = 1, 2, ..., m\}$. We can solve this problem by

$$S_T w_i = \lambda_i w_i$$

.

Fisherfaces based on FLD

Since the learning set is labeled, dimensionality reduction method should take label information into consideration. The distance between images with same label is suppose to be small. Also, the distance between different labels should be large.
Given a set of N sample images $\{x_1, x_2, ..., x_N\}$,$x_i \in \mathbb{R}^n$. And each image belongs to one of c classes $\{X_1, X_2, ..., X_c\}$. The between-class scatter matrix is defined to evaluate the distance between different classes:

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T$$

And the within-class scatter matrix is defined to evaluate the distance between images within same class:

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T$$

where μ_i is the mean image of class X_i, N_i is the number of samples in class X_i, and μ is the mean image of all samples.
If S_W is nonsingular, the optimal projection W_{opt} is chosen as the matrix with orthonormal columns which maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples.

$$\begin{aligned} W_{opt} &= \arg\max_W \frac{|W^T S_B W|}{|W^T S_W W|} \\ &= [w_1 w_2 ... w_m] \end{aligned}$$

We can solve this problem by

$$\begin{aligned} S_B w_i &= \lambda_i S_W w_i \\ S_W^{-1} S_B w_i &= \lambda_i w_i \end{aligned}$$

However, S_W is always singular according to its definition. This problem is avoided by projecting the image set to a lower dimensional space resulting within-class scatter matrix S_W nonsingular. Formally, images are projected into a N-c dimensional space using PCA and projected into c-1 dimensional space using Fisher's linear discriminant(FLD).

$$W_{opt}^T = W_{fld}^T W_{pca}^T W_{pca} = \arg\max_W |W^T S_T W|$$

$$W_{fld} = \arg\max_W \frac{|W^T W_{pca}^T S_B W_{pca} W|}{|W^T W_{pca}^T S_W W_{pca} W|}$$

Plan for next week

- Complete face recognition program.